

DOCUMENT RESUME

ED 036 750

AC 006 556

AUTHOR Martin, James L.  
TITLE Functional Path Analysis as a Multivariate Technique  
in Developing a Theory of Participation in Adult  
Education.  
PUB DATE Feb 70  
NOTE 14p.; Paper presented at the Adult Education  
Research Conference, Minneapolis, Minnesota,  
February 27-28, 1970  
EDRS PRICE MF-\$0.25 HC-\$0.80  
DESCRIPTORS Adult Education, \*Critical Path Method, \*Models,  
Participation, \*Research Methodology, \*Statistical  
Analysis

ABSTRACT

This paper reports on attempts by the author to construct a theoretical framework of adult education participation using a theory development process and the corresponding multivariate statistical techniques. Two problems are identified: the lack of theoretical framework in studying problems, and the limiting of statistical analysis to univariate techniques. The process of theory development is divided into: (a) definition of concepts; (b) establishment of variable relationships; (c) specification of causal ordering of variables; (d) identification or determination of relationships among variables within the same causal system; and (e) estimation of change. The use of regression analysis and models in the final phase is discussed, as is also functional path analysis to determine whether the proposed set of interpretations is consistent throughout. The final results will be presented in a doctoral thesis.  
(author/pt)

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE  
PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS  
STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION  
POSITION OR POLICY.

ED036750

FUNCTIONAL PATH ANALYSIS AS A MULTIVARIATE TECHNIQUE  
IN DEVELOPING A THEORY OF PARTICIPATION IN ADULT EDUCATION\*

James L. Martin

TO ERIC AND ORGANIZATIONS OPERATING  
UNDER AGREEMENTS WITH THE U.S. OFFICE OF  
EDUCATION. FURTHER REPRODUCTION OUTSIDE  
THE ERIC SYSTEM REQUIRES PERMISSION OF  
THE COPYRIGHT OWNER."

"PERMISSION TO REPRODUCE THIS  
COPYRIGHTED MATERIAL HAS BEEN GRANTED  
BY James L. Martin

Who participates in which types of adult education activities?

Why do they participate?

Adult education participation research has yielded inconclusive answers to these questions. Knox and Videbeck stated that past studies have examined only a few of the relevant variables. What is needed they suggested was a more general theory of participation. Equally restricting in my opinion are three additional limitations often used as research procedures. These are (1) the broad participant--nonparticipant category used as the dependent measure of participation; (2) the study of participation--a phenomena with apparently complex relationships--primarily with univariate statistical techniques; and (3) the lack of an explicit rationale for selecting variables. These limitations have made interpretation of results difficult if not misleading, and have hindered theory development.

Participation research in adult education has been mainly descriptive. Description is the initial step, and an essential one, in constructing a theory. However I believe it is now time researchers in adult education used this descriptive, conceptual base to build an explanatory framework which simultaneously considers all the data.

\*Paper presented at the AERA conference (Adult Education section), Minneapolis, Minnesota. February, 1970.

AC006556

Merely considering all of the variables simultaneously will not solve these four problems. In addition researchers must consider how the variables are related to each other. Participation researchers today simply do not know enough about relationships among the variates.

Adult education researchers are not alone in this quandary. Wylie(12) considered the following as "4 key flaws" of research in the area of Self Concept.

(a) the existing theories were vague, incomplete and lacking in extensive empirical support; (b) the total accumulation of substantive studies was inadequate, inconclusive and singularly disappointing . . . ; (c) the existing theories were unable to account for behaviors observed . . . ; (d) a most immediate need existed for the introduction of multivariate studies that would allow investigators to resolve systematically instead of haphazardly, many **contradictions** associated with the possible interaction of variables."

A distinction is made between theory development and theory testing or verification although the two processes are interrelated. This paper limits the discussion to theory development. Zetterberg(13, p. 101) indicated that the propositions of a theory can vary along two dimensions. The first is their degree of invariance, and the second is their generalizability or informational value. Invariant propositions with high informational value are known as laws of science. Propositions that vary between studies and have a low informational value are known as ordinary hypotheses. The problem of the researcher is to arrive at laws from these ordinary hypotheses. This means that if our aim is to arrive at laws we can proceed by accumulating findings (invariant propositions with low informational value) or we can proceed using theoretical

hypotheses (variant propositions with high informational value). Although only experience will suggest which procedure is more efficient Popper (8, p. 146) pointed out, "Science does not aim primarily, at high probabilities. It aims at high information content, well backed by experience. But a hypothesis may be very probable simply because it tells us nothing, or very little."

In theory development the interest lies with theoretical importance of the variables and not with their statistical significance. The importance of a variable is always a function of the variance measured. This is obvious in the case of regression coefficients where we are interested in the amount of change in the dependent variable produced by a given change in an independent variable.

Importance of a variable can also be assessed using a causal criterion. In the causal chain  $A \rightarrow B \rightarrow C$ ; A is considered the most important variable, although this doesn't consider the strength of the causal links. A shift of units may also vary the degree to which a particular variable is considered to be important. (1, p. 113) In assessing the relative importance among variables researchers need to know how much a given independent variable varies and what proportion of the variation in the dependent variable it explains either directly or indirectly. In effect this approach combines the quantitative criterion of importance with the causal criterion.

Theory development can be divided into five phases. First the concepts of concern to the researcher must be defined from the available literature and past research. Researchers must consider which concepts are relevant to the study, and what variables should be used to best measure these concepts. It is the responsibility of the researcher to explain the rationale underlying the selection of any particular set of variables. Second we, as researchers, must establish the two-variable relationships between the variables selected for the study. I believe that adult education participation research has been concerned primarily with this second stage. The third stage--specification--consists of specifying the causal ordering of the independent variables in the system. Statistics do not imply cause and effect. All inferences of this nature must be derived from the conceptual and theoretical interpretations of the statistical results. What is important is that a causal ordering must be specified or remain implicit to any examination of three-variable or higher order relationships that control for some of the variables. Multivariate techniques are appropriate only when the causal ordering is known or assumed. Otherwise the researcher does not know whether the data have been controlled on the dependent variables or on the independent variables.

Zetterberg (13, p.69-72) outlined the attribute-pairs among different levels of causal relationships. These are:



- 1) reversible (if X, then Y; and if Y then X) or irreversible (if X, then Y; but if Y then no conclusion about X)
- 2) deterministic (if X, then always Y) or stochastic (if X, then probably Y)
- 3) sequential (if X, then later Y) or coextensive (if X, then also Y)
- 4) sufficient (if X, then Y regardless) or contingent (if X, then Y, but only if Z)
- 5) necessary (if X, and only if X, then Y) or substitutive (if X, then Y, but if Z then also Y)

The problem of specification for theory development then is to specify both the causal ordering of the variables, and the type of causal relationships involved.

The fourth phase of theory development is identification or determining coefficients of relationships among variables within the causal system. This is the study of direct and indirect effects of each variable in the system on subsequent variables in the system. The final phase of theory development is estimation. In a structural system of equations this is the familiar problem of regression analysis with one exception. In estimation we consider the net effect on the dependent variable through both direct and indirect paths rather than assume all other independent variables are constant.

Of interest in theory development are techniques that stabilize the coefficients in the hope of giving them theoretical importance. To do this, "(researchers) . . . need either a rather tight causal scheme and a corresponding restriction on the variables considered or an arbitrary

and iron-bound choice of a set of variables. The former is clearly preferable." (Tukey p. 44) It can be shown that the correlation coefficients cannot remain the same over a wide range of situations, but it is possible that the regression coefficients might. Blalock (1, p. 50) points out that the correlations are a function of the relationship between the dependent variables and the independent variables but also reflect the degree to which other variables in the system have been brought under control. Putting it another way Blalock (2, p. 83) concluded when one's focus is directly on structure of relationships it is obvious that concern should be centered on the regression coefficients rather than correlations that merely measure goodness of fit. However in testing the adequacy of certain causal models we may make use of prediction equations involving the disappearance of either the partial regression coefficient or the correlations. In these models correlations are usually used since they are easier to deal with and interpret.

Tukey (9, p. 39) gave three situations in which the use of correlations is warranted. The first is the one described above by Blalock. Tukey has shown that in this case the correlations and regression coefficients are equal. Second, correlations should be used when exact measurement on a deterministic scale is hopeless. The third situation which warrants the use of correlations is when a researcher is dealing with specific situations. Regression only deals with the hypothetical changes which may or may not take place.

Tukey (9,p.36) classified the methods of data analysis as:

- 1) descriptive--covers only the particular populations, samples or perhaps nothing beyond the sample itself
- 2) tangential--hints or specifies what we may expect of small changes, and
- 3) functional--quite general, and is therefore the best of the three.

Bogue (3, p.4) describes these same categories as, "The distinction between 'statistical accounting for behavior', tentatively 'explaining behavior', and 'forecasting future behavior' . . . (which) must be maintained as a rule of analysis." Tukey (9, p. 41) explains the shift from the tangential to the structural class of analysis by stating that when we correct for attenuation of the regression coefficients in terms of the analysis of covariance and its relatives we pass from predictive regression to structural. Sooner or later almost all causal theory comes to deal with structural regression and structural correlation. However structural correlation equals unity and is therefore meaningless. Causation is then a guide and framework for structural analysis.

Functional path analysis is a statistical technique which uses a structural system of simultaneous regression equations to evaluate the proposed causal system. As such path analysis enters into theory development at the identification stage after the causal ordering has been specified.



Before proceeding with the discussion of path analysis I'd like to discuss how multivariate statistics can enter into each of the phases of theory development as I have outlined them.

Factor analytic techniques are useful for defining the concepts to be studied and determining their most effective measures. In the first phase of theory development factor analysis applies when several alternative variables have been measured using several instruments and the most appropriate are to be selected. Harris and Harris (5) have recently reported a strategy to determine the common factors in a set of data. This strategy involves obtaining initial solutions using Kaiser's Alpha, a uniqueness rescaling; Harris'  $R^2$ , a communality rescaling; and Joereskog's Maximum Likelihood solution. Each of these initial solutions is rotated both orthogonally and obliquely using the Kaiser normal varimax criterion. Any variable that loads higher than 0.30 absolute is considered to be a relevant variable. Common Comparable Factors are factors that have two or more of the same relevant variables on four of the six derived solutions.

In the second stage of theory development the two-variable relationships have been the most often studied in adult education participation research. The problem is that seldom have studies gone beyond this stage and established the causal relationships or established the attributes of these causal relationships.

Adult education researchers who are interested in functional analysis must be concerned with statistical techniques used for examining the theoretical order of variables in a system. These techniques should also help in distinguishing among the alternative attributes of the causal system. To theoretically specify causal order researchers need either experimental data or panel data where the same variables were measured on the same population at two different times. Linn (7) has outlined four different causal models based on different statistical procedures and the assumptions involved in each model for specifying the causal ordering among variables. In general correlational techniques can't distinguish between reversible and irreversible situations. Correlations yield only one coefficient to deal with two relationships. In this case no unique solution is mathematically possible.

Regression can be used to distinguish between these two systems and we can then answer the question; does X cause Y or does Y cause X? To specify the order of variables in a three variable system using regression as in Figure 1,

Figure 1

Time 1	Time 2
$X_1$	$X_2$
$Y_1$	$Y_2$
$Z_1$	$Z_2$

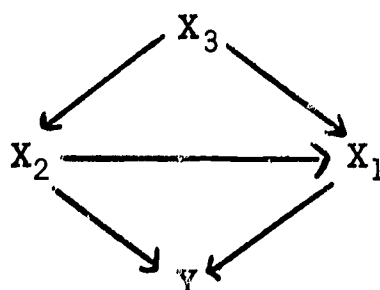
$X_2$  is regressed on  $X_1$ ,  $Y_1$ , and  $Z_1$ ;  $Y_2$  is regressed on  $X_1$ ,  $Y_1$ , and  $Z_1$ ; and  $Z_2$  is regressed on  $X_1$ ,  $Y_1$ , and  $Z_1$ .

This system of regression equations can also help determine whether Z causes both X and Y, or whether Z is an intervening variable between X and Y. If  $b_{yx} > 0.20$  then consider X causes Y, and if  $b_{xz} > 0.20$  consider Z causes X. Problems arise when the regression coefficients are small, i.e. less than 0.20. Then it becomes impossible to determine if the relationships are spurious, due to sampling error, or reflect a real, but weak relationship.

The indiscriminant use of single multiple regression equations can yield misleading results in evaluating importance. Instead the researcher needs to consider an entire set of simultaneous equations. (2, p. 866) Using results of past research and current theory the causal model is written as a set of structural equations that represent the causal processes assumed to operate among the variables under consideration. A functional path analysis of these structural equations leads to parameter estimation procedures and an evaluation of the model (6, p. 4).

Blalock (2, p. 869) supports the above argument. To represent such a causal network we need not one, but a number of separate equations, which when taken simultaneously can be used to predict changes in the dependent variable. Assuming linearity the network in Figure 2 illustrates the use of a structural system of equations in a path analysis framework used to isolate direct and indirect effects of a change in one variable on subsequent variables in the system.

Figure 2



This figure can be represented by the following set of equations:

$$Y = a_1 + b_{y1}X_1 + b_{y2}X_2 + e_1$$

$$X_1 = a_2 + b_{12}X_2 + b_{13}X_3 + e_2$$

$$X_2 = a_3 + b_{23}X_3 + e_3$$

where the  $e$ 's represent the effects of all variables not taken into consideration in the causal model. If  $X_3$  were to change there would be direct effects on  $X_1$  and  $X_2$  but only an indirect effect on  $Y$ . If  $X_2$  were to change there would be no effect on  $X_3$  but both  $X_1$  and  $Y$  would be changed. Then the total effect on  $Y$  would stem not only from the direct effect of  $X_2$  but from the change in  $X_1$  as well. In any case we could estimate the change in  $Y$  by making use of the complete set of equations. Thus, if  $X_3$  changes by one unit,  $X_2$  will change  $b_{23}$  units. The change in  $X_1$  will then be  $b_{12}b_{23} + b_{13}$ , the second term being due to  $X_3$  directly and the first to the indirect effect through  $X_2$ . We can now predict a change in  $Y$  in a similar fashion. This change will be:

$$b_{y1}(b_{12}b_{23} + b_{13}) + b_{y2}b_{23}$$

A single regression equation assuming the same causal ordering could be set up to ascertain a change in  $Y$  for a change in  $X_3$  holding both  $X_1$  and  $X_2$  constant. In such a case  $Y$  would not change. This is

both correct and misleading unless we clearly understand that we have raised a hypothetical question which in real life may be absurd.

In predictive causal models the researcher faces problems of interpretation and in ascertaining the consistency of the propositions. The problem of multiple regression in a causal framework then becomes one of substantive interpretation.

Path analysis focuses on the interpretation problem and doesn't purport to measure the cause. The purpose of path analysis is to determine whether a proposed set of interpretations is consistent throughout. Path analysis, as the name implies, is the analysis of paths within the structural diagram. Wright (11, p. 193) had defined path analysis as a technique by which the correlation between any two variables in a properly constructed diagram, e.g. Figure 2, is equal to the sum of the contributions pertaining to the path by which one may trace from one variable to another in the diagram without going back after going forward, and without passing through any variable twice in the same path. Using this definition we can trace three separate paths in Figure 2. One path runs from  $X_3$  through  $X_2$  to  $Y$ . A second path runs from  $X_3$  through  $X_1$  to  $Y$ , and the third path runs from  $X_3$  through  $X_2$  and through  $X_1$  to  $Y$ . Several paths between two variables can also be identified by following the arrows in Figure 2. The three paths from  $X_3$  to  $Y$  are three possible routes that effects of a change on  $X_3$  can follow to affect  $Y$ . Analysis of these different paths within the structural system of regression equations is path analysis.



I followed the theory development process outlined above in an attempt to establish a theoretical model for studying participation in adult education activities. My study involved a reanalysis of data on 386 young adults from rural Wisconsin. These individuals had been tested during first, sixth, ninth and twelfth grades. A mailed questionnaire was sent five years after high school graduation to ascertain participation in adult education.

The Harris factor analytic strategy was used successfully to define ten Common Comparable Factors among the 46 independent variables measured at ninth and twelfth grades. Ten separate aspects of participation were also identified from the original 25 measures using the Harris strategy. To date analysis of the two-variable relationships has failed to indicate any independent variables which have a strong relationship with any of the dependent measures.

The final results of this study will be reported in my doctoral dissertation at the University of Wisconsin, Madison, which is scheduled to be completed by June, 1970.

ERIC Clearinghouse

FEB 25 1970

on Adult Education

#### BIBLIOGRAPHY

1. Blalock, Hubert M., Causal Inferences in Nonexperimental Research. Chapel Hill, University of North Carolina Press. 1964.
2. Blalock, Hubert M., "Evaluating the Relative Importance of Variables" American Sociological Review. XXVI (December 1961) 866-874.
3. Bogue, D. J., and Harris D. L., Comparative Population and Urban Research via Multiple Regression and Covariance. Scripps Foundation, Oxford, Ohio 1954, pp. 3-18.
4. Dubin, Robert, Theory Building. Free Press, 1969.
5. Harris, M. L., and Harris, C. W., "A Factor Analytic Interpretation Strategy". Technical Report No. 120. Wisconsin Research and Development Center for Cognitive Learning, The University of Wisconsin, Madison, 1970.
6. Land, Kenneth C., "Principles of Path Analysis," in Borgatta (Ed.) Sociological Methodology. San Francisco, Jossey-Bass, 1969.
7. Linn, R. L., and Werts, C. E., "Assumptions in Making Causal Inferences from Part Correlations, Partial Correlations, and Partial Regression Coefficients." Psychological Bulletin, Vol. 72, No. 5 (November 1969), pp. 307-310.
8. Popper, Karl, "Degree of Confirmation," British Journal for the Philosophy of Science, Vol 6, 1955, p. 146.
9. Tukey, John W., "Causation, Regression, and Path Analysis," in Kempthorne (Ed.) Statistics and Mathematics in Biology, Ames, Iowa, Iowa State College Press, 1954, pp. 35-66.
10. Wright, Sewall, "The Treatment of Reciprocal Interaction, With or Without Lag, in Path Analysis." Biometrics XVI (September 1960) pp. 423-445.
11. Wright, Sewall, "Path Coefficients and Path Regressions: Alternative or Complimentary Concepts?", Biometrics XVI (June 1960) pp. 189-202.
12. Wylie, Ruth C., "The Self Concept: A Critical Survey of Pertinent Research Literature." Lincoln, University of Nebraska Press, 1961.
13. Zetterberg, Hans L., On Theory and Verification in Sociology. Totowa Bedminster Press, 1966.